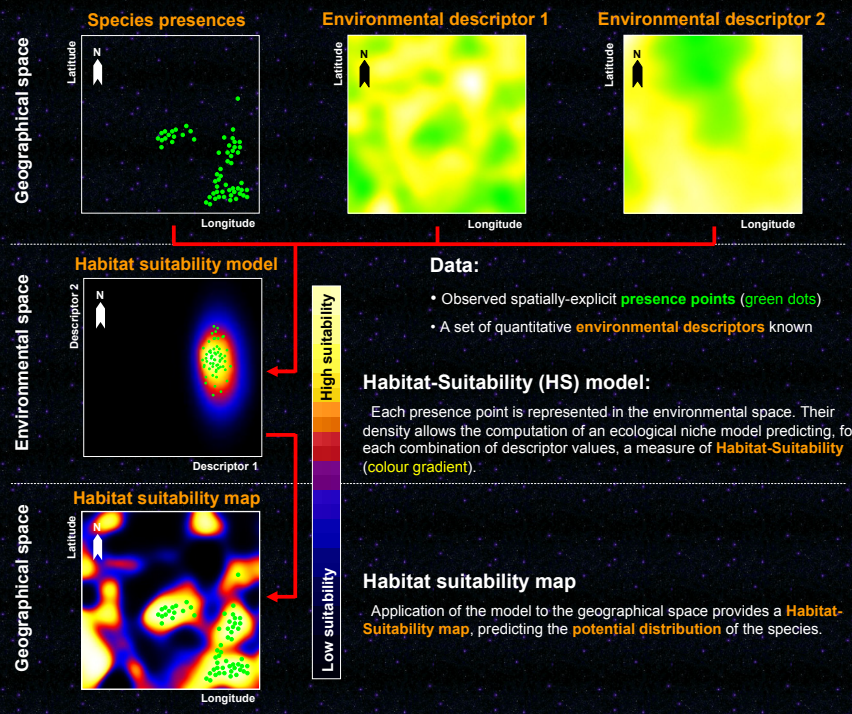
Evaluating presence-only habitat-suitability models

Hirzel A.H*., Le Lay G., Helfer V., Randin C. and Guisan A.
*E-mail: Alexandre.Hirzel@unil.ch

UNIL | Université de Lausanne
Département d'écologie et évolution

1. Habitat-suitability modelling



Data:

- Observed spatially-explicit **presence points** (green dots)
- A set of quantitative **environmental descriptors** known

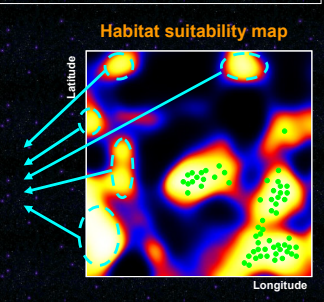
Habitat-Suitability (HS) model:

Each presence point is represented in the environmental space. Their density allows the computation of an ecological niche model predicting, for each combination of descriptor values, a measure of **Habitat-Suitability** (colour gradient).

Habitat suitability map

Application of the model to the geographical space provides a **Habitat-Suitability map**, predicting the **potential distribution** of the species.

2. Evaluation problems



Principles:

To evaluate the predictive power of an HS map consists in general to assess whether it predicts high suitability where the species is observed, and low suitability where it is absent.

Presence-only modelling:

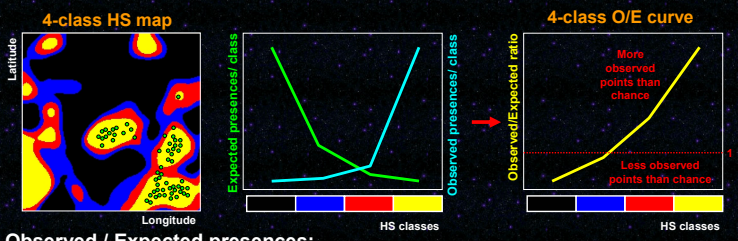
In the case of **presence-only HS modelling** however, **absences are unreliable**. A species may be absent from suitable habitat for a number of reasons (temporary local extinction, barriers to dispersal, invasive species, disturbances, cryptic species, sampling bias, etc.)

Therefore, **absences may not be used to evaluate presence-only HS models**.

Problems:

- As they all (e.g. Kappa, ROC AUC) rely on presences and absences, **current evaluators are useless** for presence-only models.
- Using only presences is difficult because nothing counterbalances them in the evaluation. Therefore, the best model would be the one predicting suitable habitat everywhere.

3. Solution: the Continuous Boyce index

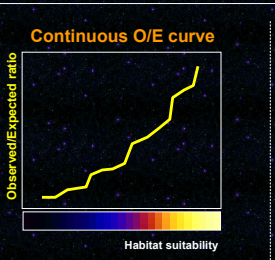


Observed / Expected presences:

The HS map is reclassified into **four classes** (for instance. Note that classical evaluators reclassify into 2 classes, suitable/unsuitable). For each class, compute:

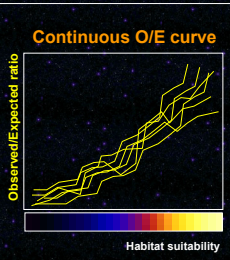
- The **expected proportion** of presences per class if the species was distributed at random (E_i =Class$_i$ area/Total area).
- The **observed proportion** of presences per class (O_i= Class$_i$ presence count / Total count)
- The evaluator F, given by:

$$F_i = \frac{O_i}{E_i}$$



Continuous O/E curve:

Instead of a fixed number of classes, we can compute F=O/E over a **moving class** of fixed width (say 20% the total HS range). This provides a **continuous O/E curve** that does not depend on the number of classes or of class boundaries.

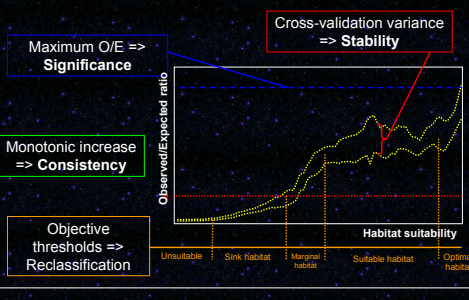


Cross-validation:

The presence points are **partitioned into 10 parts**. The HS model is calibrated with nine of them and evaluated on the last one. This is repeated 10 times, switching the evaluation partition.

This provides 10 O/E curves.

4. Interpretation & Conclusion



The cross-validated O/E curves assess the model **consistency**, **stability**, and **significance**.

It allows to define objective thresholds for **reclassification**, providing more honest and relevant HS maps.

Tests with 114 plants showed the continuous Boyce index to be **consistent** with **Kappa** and **ROC AUC**.

5. References

- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199**(2), 142-52.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**(2-3). 281-300.
- Implemented into **Biomapper 3.3**: **www.unil.ch/biomapper**